Extremality Properties for the Basic Polarization Transformations

Mine Alsan

Information Theory Laboratory

Ecole Polytechnique Fédérale de Lausanne

CH-1015 Lausanne, Switzerland

Email: mine.alsan@epfl.ch

Abstract

We study the extremality of the BEC and the BSC for Gallager's reliability function E_0 evaluated under the uniform input distribution for binary input DMCs from the aspect of channel polarization. In particular, we show that amongst all B-DMCs of a given $E_0(\rho)$ value, for a fixed $\rho \geq 0$, the BEC and BSC are extremal in the evolution of E_0 under the one-step polarization transformations.

Index Terms

Channel reliability function, channel polarization, extremal channels

I. Introduction

While the capacity of a memoryless channel W gives the largest rate that may be communicated reliably across it, the reliability function $E(R, W)$ provides a finer measure on the quality of the channel: for any rate R less than channel capacity, it is possible to find a sequence of codes of increasing blocklength, each of which of rate at least R, and whose block error probability decays exponentially to zero in the blocklength — $E(R, W)$ is the largest possible rate of this decay.

Gallager classical treatise [1] gives a lower bound to $E(R, W)$, the random coding exponent $E_r(R, W)$ in the form $E_r(R, W) = \max_{\rho \in [0,1]} E_0(\rho, W) - \rho R$. Remarkably, this lower bound is tight for rates above the critical rate $E_0'(1, W)$. The function $E_0(\rho, W)$ that appears as an auxiliary function on the road to deriving $E_r(R, W)$ turns out to be of independent interest in its own right. In particular, $E_0(\rho, W)/\rho$ is the largest rate for which a sequential decoder can operate while keeping the ρ-th moment of the decoder's computation effort per symbol bounded.

In [2], we investigated the extremal properties of $E_0(\rho, W)$ evaluated under the uniform input distribution for the class of binary input channels. We have shown that among all such channels with a given value of $E_0(\rho_1, W)$, for $\rho_1 \in [0, 1]$, the binary erasure channel (BEC) and the binary symmetric channel (BSC) distinguish themselves in certain ways: they have, respectively, the largest and smallest value of $E_0'(\rho_2, W)$ for any $\rho_2 \in [\rho_1, 1]$. Furthermore, we showed that amongst channels W with a given value of $E_0(\rho, W)$ for a given $\rho \in [0, 1]$, the BEC and BSC are the most and least polarizing under Arıkan's polar transformations in the sense that their polar transforms W^+ and W^- have the largest and smallest difference in their E_0 values.

In this paper, we extend the result related to the BEC and BSC being extremal for Arıkan's polarization transforms to the region where $\rho \geq 0$. In his award winning paper [3], Arıkan describes two synthetic channels W^+, and W^- which can be obtained from two independent copies of W. It is well known (proved as a corollary to extremes of information combining) that among all channels W with a given symmetric capacity $I(W)$, the BEC and BSC polarize most and least in the sense of having the largest and smallest difference between $I(W^+)$ and $I(W^-)$. We report a more general conclusion: amongst all channels W with a given value of $E_0(\rho, W)$, the BEC and BSC polarize most and least in the sense of having the largest difference between $E_0(\rho, W^+)$ and $E_0(\rho, W^-)$ whenever $\rho \in [0, 1] \cup [2, \infty]$. On the other hand, for $\forall \rho \in [1, 2]$, we show that the BEC maximizes, and the BSC minimizes the E_0 values obtained after both applying the W^+, or the W^- transformations.

A. Definitions

Given a binary input channel W, let $E_0(\rho, W)$ denote "Gallager's E_0" [1, p. 138] evaluated for the uniform input distribution:

$$E_0(\rho, W) = -\log \sum_{y \in \mathcal{Y}} \left[\frac{1}{2} W(y \mid 0)^{\frac{1}{1+\rho}} + \frac{1}{2} W(y \mid 1)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \tag{1}$$

Theorem 5.6.3 in [1] summarizes the properties of $E_0(\rho, W)$ with respect to the variable ρ. For $\rho \geq 0$, $E_0(\rho, W)$ is a positive, concave increasing function in ρ. Moreover, the symmetric capacity $I(W)$ of the channel can be derived from $E_0(\rho, W)$ by

$$\lim_{\rho \to 0} \frac{E_0(\rho, W)}{\rho} = \frac{\partial}{\partial \rho} E_0(\rho, W)\Big|_{\rho=0} = I(W) \tag{2}$$

and the Bhattacharyya parameter $Z(W)$ from the cut-off rate as

$$E_0(1, W) = \log \frac{2}{1 + Z(W)}. \tag{3}$$

The next lemma due to Telatar and Arıkan [4] introduces a useful representation for the $E_0(\rho, W)$ parameter.

Lemma 1: [4] Given a symmetric B-DMC W, and a fixed $\rho \in [0, 1]$, there exist a random variable Z taking values in the $[0, 1]$ interval such that

$$E_0(\rho, W) = -\log \mathbb{E}\left[g(\rho, Z)\right] \tag{4}$$

where

$$g(\rho, z) = \left(\frac{1}{2}(1 + z)^{\frac{1}{1+\rho}} + \frac{1}{2}(1 - z)^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{5}$$

Moreover, the random variable Z_{BEC} of a binary erasure channel is $\{0, 1\}$ valued. The random variable Z_{BSC} of a binary symmetric channel is a constant z_{BSC}.

Proof: Recall $E_0(\rho, W) = -\log \sum_y \left[\frac{1}{2}W(y \mid 0)^{\frac{1}{1+\rho}} + \frac{1}{2}W(y \mid 1)^{\frac{1}{1+\rho}}\right]^{1+\rho}$. Define

$$q_W(y) = \frac{W(y \mid 0) + W(y \mid 1)}{2} \quad \text{and} \quad \Delta_W(y) = \frac{W(y \mid 0) - W(y \mid 1)}{W(y \mid 0) + W(y \mid 1)} \tag{6}$$

so that $W(y \mid 0) = q_W(y)[1 + \Delta_W(y)]$ and $W(y \mid 1) = q_W(y)[1 - \Delta_W(y)]$. Then, one can define the random variable $Z = |\Delta_W(Y)| \in [0, 1]$ where Y has the probability distribution $q_W(y)$, and obtain (4) by simple manipulations. The claims about Z_{BEC}, and Z_{BSC} are verified easily from (6). ∎

II. EXTREMALITY RESULTS FOR THE POLARIZATION TRANSFORMATIONS

A. Basic Polarization Transformations

In [3], a low complexity code construction that achieves the symmetric capacity of B-DMCs is given based on the recursive application of two basic channel transformations. These transforms, usually refered as the minus and plus transformations, synthesize two new channels by combining

two independent copies of a given channel. The transition probabilities of the new channels are defined in terms of the initial one by the definitions given in [3, Eqs. (19), (20)].

Instead of identical copies of a given channel, we propose to combine two independent copies of different B-DMCs in a similar way. We denote by $W_{1,2}^- : \mathcal{X} \to \mathcal{Y}^2$ and $W_{1,2}^+ : \mathcal{X} \to \mathcal{Y}^2 \times \mathcal{X}$ the synthesized channels obtained by combining independent copies of the channels W_1 and W_2. In this case, the transition probabilities can be defined by

$$W_{1,2}^-(y_1 y_2 \mid u_1) = \sum_{u_2 \in \mathcal{X}} \frac{1}{2} W_1(y_1 \mid u_1 \oplus u_2) W_2(y_2 \mid u_2) \tag{7}$$

$$W_{1,2}^+(y_1 y_2 u_1 \mid u_2) = \frac{1}{2} W_1(y_1 \mid u_1 \oplus u_2) W_2(y_2 \mid u_2). \tag{8}$$

The following two lemmas express the E_0 parameter of the synthesized channels $W_{1,2}^-$, and $W_{1,2}^+$ in terms of the representation given in Lemma 1, relating them to the E_0 parameters of the channels W_1 and W_2.

Lemma 2: Given two B-DMCs W_1, W_2, and $\rho \geq 0$, let Z_1 and Z_2 be independent RVs such that

$$E_0(\rho, W_1) = -\log \mathbb{E}\left[g(\rho, Z_1)\right] \quad \text{and} \quad E_0(\rho, W_2) = -\log \mathbb{E}\left[g(\rho, Z_2)\right]$$

hold as defined in Lemma 1. Then,

$$E_0(\rho, W_{1,2}^-) = -\log \mathbb{E}\left[g(\rho, Z_1 Z_2)\right] \tag{9}$$

where $g(\rho, z)$ is given by (5).

Proof: From the definition of the channel $W_{1,2}^-$ in (7), we can write

$$E_0(\rho, W_{1,2}^-) = -\log \sum_{y_1, y_2} \left[\frac{1}{2} W_{1,2}^-(y_1, y_2 \mid 0)^{\frac{1}{1+\rho}} + \frac{1}{2} W_{1,2}^-(y_1, y_2 \mid 1)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

$$= -\log \sum_{y_1, y_2} \left[\frac{1}{2} \left(\frac{1}{2} W_1(y_1 \mid 0) W_2(y_2 \mid 0) + \frac{1}{2} W_1(y_1 \mid 1) W_2(y_2 \mid 1) \right)^{\frac{1}{1+\rho}} \right.$$

$$\left. + \frac{1}{2} \left(\frac{1}{2} W_1(y_1 \mid 1) W_2(y_2 \mid 0) + \frac{1}{2} W_1(y_1 \mid 0) W_2(y_2 \mid 1) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

$$= -\log \sum_{y_1 y_2} \left[\frac{1}{2} \left(\frac{1}{2} \right)^{\frac{1}{1+\rho}} q_{W_1}(y_1)^{\frac{1}{1+\rho}} q_{W_2}(y_2)^{\frac{1}{1+\rho}} \right.$$

$$\left((1 + \Delta_{W_1}(y_1)) (1 + \Delta_{W_2}(y_2)) + (1 - \Delta_{W_1}(y_1)) (1 - \Delta_{W_2}(y_2)) \right)^{\frac{1}{1+\rho}}$$

$$\left. + \left((1 - \Delta_{W_1}(y_1)) (1 + \Delta_{W_2}(y_2)) + (1 + \Delta_{W_1}(y_1)) (1 - \Delta_{W_2}(y_2)) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

$$= -\log \sum_{y_1 y_2} q(y_1)\, q(y_2)\, \left[\frac{1}{2} \left(1 + \Delta_{W_1}(y_1) \Delta_{W_2}(y_2) \right)^{\frac{1}{1+\rho}} + \frac{1}{2} \left(1 - \Delta_{W_1}(y_1) \Delta_{W_2}(y_2) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

where we used the definitions in (6). We can now define $Z_1 = |\Delta_{W_1}(Y_1)|$ and $Z_2 = |\Delta_{W_2}(Y_2)|$ where Y_1 and Y_2 are independent random variables with distribution q_{W_1} and q_{W_2}, respectively. From this construction, the lemma follows. ∎

Lemma 3: Given two B-DMCs W_1, W_2, and $\rho \geq 0$, let Z_1 and Z_2 be as in Lemma 2. Then,

$$E_0(\rho, W_{1,2}^+) = -\log \mathbb{E} \left[\frac{1}{2}(1 + Z_1 Z_2)\ g\left(\rho, \frac{Z_1 + Z_2}{1 + Z_1 Z_2}\right) + \frac{1}{2}(1 - Z_1 Z_2)\ g\left(\rho, \frac{Z_1 - Z_2}{1 - Z_1 Z_2}\right) \right] \tag{10}$$

where $g(\rho, z)$ is given by (5).

Proof: From the definition of channel W^+ in (8), we can write

$$E_0(\rho, W^+_{1,2})$$

$$= -\log \sum_{y_1,y_2,u} \left[\frac{1}{2} W^+_{1,2}(y_1, y_2, u \mid 0)^{\frac{1}{1+\rho}} + \frac{1}{2} W^+_{1,2}(y_1, y_2, u \mid 1)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

$$= -\log \sum_{y_1,y_2,u} \left[\frac{1}{2} \left(\frac{1}{2} W_1(y_1 \mid u) W_2(y_2 \mid 0) \right)^{\frac{1}{1+\rho}} + \frac{1}{2} \left(\frac{1}{2} W_1(y_1 \mid u \oplus 1) W_2(y_2 \mid 1) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

$$= -\log \sum_{y_1,y_2} \left(\left[\frac{1}{2} \left(\frac{1}{2} W_1(y_1 \mid 0) W_2(y_2 \mid 0) \right)^{\frac{1}{1+\rho}} + \frac{1}{2} \left(\frac{1}{2} W_1(y_1 \mid 1) W_2(y_2 \mid 1) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right.$$

$$\left. + \left[\frac{1}{2} \left(\frac{1}{2} W_1(y_1 \mid 1) W_2(y_2 \mid 0) \right)^{\frac{1}{1+\rho}} + \frac{1}{2} \left(\frac{1}{2} W_1(y_1 \mid 0) W_2(y_2 \mid 1) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right).$$

Using (6), we have

$$E_0(\rho, W^+_{1,2})$$

$$= -\log \sum_{y_1 y_2} \frac{1}{2} \, q_{W_1}(y_1) \, q_{W_2}(y_2)$$

$$\left(\left[\left((1 + \Delta_{W_1}(y_1)) \, (1 + \Delta_{W_2}(y_2)) \right)^{\frac{1}{1+\rho}} + \left((1 - \Delta_{W_1}(y_1)) \, (1 - \Delta_{W_2}(y_2)) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right.$$

$$\left. + \left[\left((1 - \Delta_{W_1}(y_1)) \, (1 + \Delta_{W_2}(y_2)) \right)^{\frac{1}{1+\rho}} + \left((1 - \Delta_{W_1}(y_1)) \, (1 + \Delta_{W_2}(y_2)) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right)$$

$$= -\log \left(\sum_{y_1 y_2} \frac{1}{2} \, q_{W_1}(y_1) \, q_{W_2}(y_2) \, \left(1 + \Delta_{W_1}(y_1) \Delta_{W_2}(y_2) \right) \right.$$

$$\left[\frac{1}{2} \left(1 + \frac{\Delta_{W_1}(y_1) + \Delta_{W_2}(y_2)}{1 + \Delta_{W_1}(y_1) \Delta_{W_2}(y_2)} \right)^{\frac{1}{1+\rho}} + \frac{1}{2} \left(1 - \frac{\Delta_{W_1}(y_1) + \Delta_{W_2}(y_2)}{1 + \Delta_{W_1}(y_1) \Delta_{W_2}(y_2)} \right)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

$$+ \sum_{y_1 y_2} \frac{1}{2} \, q_{W_1}(y_1) \, q_{W_2}(y_2) \, \left(1 - \Delta_{W_1}(y_1) \Delta_{W_2}(y_2) \right)$$

$$\left[\frac{1}{2} \left(1 + \frac{\Delta_{W_1}(y_1) - \Delta_{W_2}(y_2)}{1 - \Delta_{W_1}(y_1) \Delta_{W_2}(y_2)} \right)^{\frac{1}{1+\rho}} + \frac{1}{2} \left(1 - \frac{\Delta_{W_1}(y_1) - \Delta_{W_2}(y_2)}{1 - \Delta_{W_1}(y_1) \Delta_{W_2}(y_2)} \right)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right)$$

$$= -\log \left(\sum_{y_1 y_2} \frac{1}{2} \, q_{W_1}(y_1) \, q_{W_2}(y_2) \, \left(1 + \Delta_{W_1}(y_1) \Delta_{W_2}(y_2) \right) \, g\!\left(\rho, \frac{\Delta_{W_1}(y_1) + \Delta_{W_2}(y_2)}{1 + \Delta_{W_1}(y_1) \Delta_{W_2}(y_2)} \right) \right.$$

$$\left. + \sum_{y_1 y_2} \frac{1}{2} \, q_{W_1}(y_1) \, q_{W_2}(y_2) \, \left(1 - \Delta_{W_1}(y_1) \Delta_{W_2}(y_2) \right) \, g\!\left(\rho, \frac{\Delta_{W_1}(y_1) - \Delta_{W_2}(y_2)}{1 - \Delta_{W_1}(y_1) \Delta_{W_2}(y_2)} \right) \right)$$

where $g(\rho, z)$ is defined in (5).

Similar to the $E_0(\rho, W_{1,2}^-)$ case, we define $Z_1 = |\Delta_{W_1}(Y_1)|$ and $Z_2 = |\Delta_{W_2}(Y_2)|$ where Y_1 and Y_2 are independent random variables with distributions q_{W_1} and q_{W_2}, respectively. However, we should check whether this construction is equivalent to the above equation. We note that $\Delta \in [-1, 1]$. When $\Delta_{W_1}(y_1)$ and $\Delta_{W_2}(y_2)$ are of the same sign, we can easily see (noting that $g(\rho, z)$ is symmetric about $z = 0$) that

$$\left(1 + \Delta_{W_1}(y_1)\Delta_{W_2}(y_2)\right) \ g\left(\rho, \frac{\Delta_{W_1}(y_1) + \Delta_{W_2}(y_2)}{1 + \Delta_{W_1}(y_1)\Delta_{W_2}(y_2)}\right) = \left(1 + Z_1 Z_2\right) \ g\left(\rho, \frac{Z_1 + Z_2}{1 + Z_1 Z_2}\right)$$

$$\left(1 - \Delta_{W_1}(y_1)\Delta_{W_2}(y_2)\right) \ g\left(\rho, \frac{\Delta_{W_1}(y_1) - \Delta_{W_2}(y_2)}{1 - \Delta_{W_1}(y_1)\Delta_{W_2}(y_2)}\right) = \left(1 - Z_1 Z_2\right) \ g\left(\rho, \frac{Z_1 - Z_2}{1 - Z_1 Z_2}\right)$$

When $\Delta_W(y_1)$ and $\Delta_W(y_2)$ are of the opposite sign, we note that

$$\left(1 + \Delta_{W_1}(y_1)\Delta_{W_2}(y_2)\right) \ g\left(\rho, \frac{\Delta_{W_1}(y_1) + \Delta_{W_2}(y_2)}{1 + \Delta_{W_1}(y_1)\Delta_{W_2}(y_2)}\right) = \left(1 - Z_1 Z_2\right) \ g\left(\rho, \frac{Z_1 - Z_2}{1 - Z_1 Z_2}\right)$$

$$\left(1 - \Delta_{W_1}(y_1)\Delta_{W_2}(y_2)\right) \ g\left(\rho, \frac{\Delta_{W_1}(y_1) - \Delta_{W_2}(y_2)}{1 - \Delta_{W_1}(y_1)\Delta_{W_2}(y_2)}\right) = \left(1 + Z_1 Z_2\right) \ g\left(\rho, \frac{Z_1 + Z_2}{1 + Z_1 Z_2}\right)$$

Since we are interested in the sum of the above two parts, we can see that the construction we propose is still equivalent. This concludes the proof. ∎

Remark 1: By the symmetry of the RVs Z_1 and Z_2, we have $E_0(\rho, W_{1,2}^\pm) = E_0(\rho, W_{2,1}^\pm)$.

Lemma 4: The channels $W_{1,2}^-$, W_1, W_2, and $W_{1,2}^+$ satisfy the following ordering:

$$E_0(\rho, W_{1,2}^-) \leq E_0(\rho, W_1) \leq E_0(\rho, W_{1,2}^+), \tag{11}$$

$$E_0(\rho, W_{1,2}^-) \leq E_0(\rho, W_2) \leq E_0(\rho, W_{1,2}^+).$$

Proof: We only show the inequalities in (11) for the channel W_1. The proof for the channel W_2 follows from Remark 1. By Lemmas 1, 2, and 3 the inequalities in (11) are equivalent to

$$\mathbb{E}\left[\frac{1}{2}(1 + Z_1 Z_2) \ g\left(\rho, \frac{Z_1 + Z_2}{1 + Z_1 Z_2}\right) + \frac{1}{2}(1 - Z_1 Z_2) \ g\left(\rho, \frac{Z_1 - Z_2}{1 - Z_1 Z_2}\right)\right] \leq \mathbb{E}\left[g(\rho, Z_1)\right], \tag{12}$$

$$\mathbb{E}\left[g(\rho, Z_1)\right] \leq \mathbb{E}\left[g(\rho, Z_1 Z_2)\right]. \tag{13}$$

By Lemma 7, the function $g(\rho, z)$ is non-increasing in the variable z when $\rho \geq 0$. Hence, the second inequality in (13) holds. On the other side, note that for any realizations z_1 and z_2, the

factors $\frac{1}{2}(1 + z_1 z_2)$, and $\frac{1}{2}(1 - z_1 z_2)$ form a distribution. As the function $g(\rho, z)$ is concave in z by Lemma 7, we can apply Jensen's inequality to obtain

$$\frac{1}{2}(1+z_1 z_2) \; g\Big(\rho, \frac{z_1 + z_2}{1 + z_1 z_2}\Big) + \frac{1}{2}(1-z_1 z_2) \; g\Big(\rho, \frac{z_1 - z_2}{1 - z_1 z_2}\Big) \leq g\Big(\rho, \frac{z_1 + z_2}{2} + \frac{z_1 - z_2}{2}\Big) = g(\rho, z_1).$$

Taking the expectation of both sides, we get the first inequality in (12). ∎

Remark 2: In [5] it is shown that the channels W_1, W_2, $W_{1,2}^-$, and $W_{1,2}^+$ satisfy the relationship:

$$E_0(\rho, W_{1,2}^+) + E_0(\rho, W_{1,2}^-) \geq E_0(\rho, W_1) + E_0(\rho, W_2), \quad \forall \rho \geq 0.$$

B. Extremality for the Basic Channel Transformations

Theorem 1: Given two B-DMCs W_1, and W_2, for any fixed value of $\rho \geq 0$, we define two binary symmetric channels W_{BSC}, and $W_{\overline{\mathsf{BSC}}}$, and two binary erasure channels W_{BEC}, and $W_{\overline{\mathsf{BEC}}}$ through the equalities

$$E_0(\rho, W_1) = E_0(\rho, W_{\mathsf{BEC}}) = E_0(\rho, W_{\mathsf{BSC}}), \tag{14}$$

$$E_0(\rho, W_2) = E_0(\rho, W_{\overline{\mathsf{BEC}}}) = E_0(\rho, W_{\overline{\mathsf{BSC}}}). \tag{15}$$

Then for the $W_{1,2}^-$ polar transformation, we have

$$E_0(\rho, W_{\mathsf{BEC},\overline{\mathsf{BEC}}}^-) \leq E_0(\rho, W_{1,2}^-) \leq E_0(\rho, W_{\mathsf{BSC},\overline{\mathsf{BSC}}}^-) \quad \forall \rho \geq 0. \tag{16}$$

For the W^+ polar transformation, we have

$$E_0(\rho, W_{\mathsf{BSC},\overline{\mathsf{BSC}}}^+) \leq E_0(\rho, W_{1,2}^+) \leq E_0(\rho, W_{\mathsf{BEC},\overline{\mathsf{BEC}}}^+) \quad \forall \rho \in [0, 1] \cup [2, \infty], \tag{17}$$

$$E_0(\rho, W_{\mathsf{BEC},\overline{\mathsf{BEC}}}^+) \leq E_0(\rho, W_{1,2}^+) \leq E_0(\rho, W_{\mathsf{BSC},\overline{\mathsf{BSC}}}^+) \quad \forall \rho \in [1, 2]. \tag{18}$$

Proof: We start to show the result for the minus transformation given in Equation (16). This proof relies on the convexity result stated in the next lemma. The proof of the lemma is given in Appendix A.

Lemma 5: For any $z \in [0, 1]$, and $\rho \geq 0$, the function $F_{z,\rho}(t) : [2^{-\rho}, 1] \to [g(\rho, z), 1]$ defined as

$$F_{z,\rho}(t) = g(\rho, zg^{-1}(\rho, t)) \tag{19}$$

where $g^{-1}(\rho, t)$ denotes the inverse of the function g with respect to its second argument, is convex with respect to the variable t.

From Lemmas 1, and 2, we know that

$$\exp\{-E_0(\rho, W_1)\} = \mathbb{E}[g(\rho, Z_1)]$$

$$\exp\{-E_0(\rho, W_2)\} = \mathbb{E}[g(\rho, Z_2)]$$

$$\exp\{-E_0(\rho, W_{1,2}^-)\} = \mathbb{E}[g(\rho, Z_1 Z_2)]$$

where Z_1 and Z_2 are independent random variables. We also know $Z_{\mathsf{BSC}} = z_{\mathsf{BSC}}$, $Z_{\overline{\mathsf{BSC}}} = z_{\overline{\mathsf{BSC}}}$ and $Z_{\mathsf{BEC}}, Z_{\overline{\mathsf{BEC}}} \in \{0, 1\}$. Hence,

$$\exp\{-E_0(\rho, W_{\mathsf{BSC},\overline{\mathsf{BSC}}}^-)\} = g(\rho, z_{\mathsf{BSC}} z_{\overline{\mathsf{BSC}}}).$$

Given $E_0(\rho, W_1) = E_0(\rho, W_{\mathsf{BSC}})$, and $E_0(\rho, W_2) = E_0(\rho, W_{\overline{\mathsf{BSC}}})$ we also have

$$\mathbb{E}\left[g(\rho, Z_1)\right] = g(\rho, z_{\mathsf{BSC}}),$$

$$\mathbb{E}\left[g(\rho, Z_2)\right] = g(\rho, z_{\overline{\mathsf{BSC}}}).$$

Therefore, using Jensen's inequality we obtain

$$\begin{aligned}
\exp\{-E_0(\rho, W_{1,2}^-)\} &= \mathbb{E}_{Z_1}[\mathbb{E}_{Z_2}[F_{z_1,\rho}\left(g(\rho, Z_2)\right) \mid Z_1 = z_1]] \\
&\geq \mathbb{E}_{Z_1}[F_{Z_1,\rho}\left(\mathbb{E}_{Z_2}[g(\rho, Z_2)]\right)] \\
&= \mathbb{E}_{Z_1}[F_{Z_1,\rho}\left(g(\rho, z_{\overline{\mathsf{BSC}}})\right)] \\
&\stackrel{(1)}{=} \mathbb{E}_{Z_1}\left[F_{z_{\overline{\mathsf{BSC}}},\rho}\left(g(\rho, Z_1)\right)\right] \\
&\geq F_{z_{\overline{\mathsf{BSC}}},\rho}\left(\mathbb{E}_{Z_1}[g(\rho, Z_1)]\right) \\
&= F_{z_{\overline{\mathsf{BSC}}},\rho}\left(g(\rho, z_{\mathsf{BSC}})\right) \\
&= \exp\{-E_0(\rho, W_{\mathsf{BSC},\overline{\mathsf{BSC}}}^-)\}
\end{aligned}$$

where (1) follows by symmetry of the variables Z_1 and $z_{\overline{\mathsf{BSC}}}$.

Let ϵ, and $\overline{\epsilon}$ be the erasure probabilities of W_{BEC}, and $W_{\overline{\mathsf{BEC}}}$, respectively. Then, we have $P(Z_{\mathsf{BEC}} = 0) = \epsilon$, $P(Z_{\overline{\mathsf{BEC}}} = 0) = \overline{\epsilon}$, and

$$\exp\{-E_0(\rho, W_{\mathsf{BEC}})\} = P(Z_{\mathsf{BEC}} = 0)(1 - 2^{-\rho}) + 2^{-\rho},$$

$$\exp\{-E_0(\rho, W_{\overline{\mathsf{BEC}}})\} = P(Z_{\overline{\mathsf{BEC}}} = 0)(1 - 2^{-\rho}) + 2^{-\rho}.$$

The channel $W_{1,2}^-$ is a BEC with erasure probability $\epsilon + \overline{\epsilon} - \epsilon\overline{\epsilon}$, hence we get

$$\exp\{-E_0(\rho, W_{\mathsf{BEC},\overline{\mathsf{BEC}}}^-)\} = [P(Z_{\mathsf{BEC}} = 0) + P(Z_{\overline{\mathsf{BEC}}} = 0) - P(Z_{\mathsf{BEC}} = 0)P(Z_{\overline{\mathsf{BEC}}} = 0)] (1 - 2^{-\rho}) + 2^{-\rho}.$$

Therefore, given $E_0(\rho, W_1) = E_0(\rho, W_{\mathsf{BEC}})$, and $E_0(\rho, W_2) = E_0(\rho, W_{\overline{\mathsf{BEC}}})$, we have

$$\mathbb{E}\left[g(\rho, Z_1)\right] = \mathbb{E}\left[g(\rho, Z_{\mathsf{BEC}})\right] = P(Z_{\mathsf{BEC}} = 0)(1 - 2^{-\rho}) + 2^{-\rho},$$

$$\mathbb{E}\left[g(\rho, Z_2)\right] = \mathbb{E}\left[g(\rho, Z_{\overline{\mathsf{BEC}}})\right] = P(Z_{\overline{\mathsf{BEC}}} = 0)(1 - 2^{-\rho}) + 2^{-\rho}.$$

Due to convexity, we also know the following inequality holds:

$$F_{z,\rho}(t) \le (1-t)F_{z,\rho}(0) + tF_{z,\rho}(1) = 1 + \frac{g(\rho, z) - 1}{2^{-\rho} - 1}(t - 1).$$

Therefore,

$$\exp\{-E_0(\rho, W_{1,2}^-)\} \tag{20}$$

$$= \mathbb{E}_{Z_1}[\mathbb{E}_{Z_2}[F_{z_1,\rho}\left(g(\rho, Z_2)\right) \mid Z_1 = z_1]]$$

$$\le \mathbb{E}_{Z_1}\left[1 + \frac{g(\rho, Z_1) - 1}{2^{-\rho} - 1}(\mathbb{E}_{Z_2}[g(\rho, Z_2)] - 1)\right]$$

$$= 1 + \frac{\mathbb{E}_{Z_1}[g(\rho, Z_1)] - 1}{2^{-\rho} - 1}(\mathbb{E}_{Z_2}[g(\rho, Z_2)] - 1)$$

$$= 1 + \frac{[P(Z_{\mathsf{BEC}} = 0)(1 - 2^{-\rho}) + 2^{-\rho} - 1][P(Z_{\overline{\mathsf{BEC}}} = 0)(1 - 2^{-\rho}) + 2^{-\rho} - 1]}{2^{-\rho} - 1}$$

$$= 1 - P(Z_{\mathsf{BEC}} = 0)P(Z_{\overline{\mathsf{BEC}}} = 0)(1 - 2^{-\rho}) + \left(P(Z_{\mathsf{BEC}} = 0) + P(Z_{\overline{\mathsf{BEC}}} = 0)\right)(1 - 2^{-\rho}) + 2^{-\rho} - 1$$

$$= \left[P(Z_{\mathsf{BEC}} = 0) + P(Z_{\overline{\mathsf{BEC}}} = 0) - P(Z_{\mathsf{BEC}} = 0)P(Z_{\overline{\mathsf{BEC}}} = 0)\right](1 - 2^{-\rho}) + 2^{-\rho}$$

$$= \exp\{-E_0(\rho, W_{\mathsf{BEC},\overline{\mathsf{BEC}}}^-)\}.$$

This concludes the proof for the minus transformation. Now, we sketch the proof of the extremality property for the plus transformation. We define the function $h(\rho, z_1, z_2)$ as

$$h(\rho, z_1, z_2) = \frac{1}{2}\left(1 + z_1 z_2\right)g\left(\rho, \frac{z_1 + z_2}{1 + z_1 z_2}\right) + \frac{1}{2}\left(1 - z_1 z_2\right)g\left(\rho, \frac{z_1 - z_2}{1 - z_1 z_2}\right) \tag{21}$$

where $z_1, z_2 \in [0, 1]$, and $\rho \ge 0$. Note that $h(\rho, z_1, z_2)$ is symmetric in the variables z_1, and z_2. The proof relies on the convexity result stated in the next lemma. The proof of the lemma is given in Appendix B.

Lemma 6: [6] For any $z \in [0, 1]$, and $\rho \ge 0$, the function $H_{z,\rho}(t) : [2^{-\rho}, 1] \to [2^{-\rho}, g(\rho, z)]$ defined as

$$H_{z,\rho}(t) = h(\rho, g^{-1}(\rho, t), z)$$

is concave with respect to the variable t when $\rho \in [0, 1] \cup [2, \infty]$, and convex when $\rho \in [1, 2]$.

The proof of the theorem for the plus transformation can be completed following similar steps to the minus case. By Lemma 3, we have

$$\mathbb{E}\left[h(\rho, Z_1, Z_2)\right] = \exp\{-E_0(\rho, W_{1,2}^+)\}.$$

We define the random variables

$$T_1 = g(\rho, Z_1) \quad \text{and} \quad T_2 = g(\rho, Z_2).$$

Then, using the concavity of the function $H_{z,\rho}(t)$ with respect to t for fixed values of $\rho \in [0,1] \cup [2,\infty]$, and $z \in [0,1]$, we obtain the inequalities in (17):

$$\exp\{-E_0(\rho, W_{1,2}^+)\} = \mathbb{E}\left[H_{g^{-1}(\rho,T_2),\rho}(T_1)\right] \leq h(\rho, z_{\mathsf{BSC}}, z_{\overline{\mathsf{BSC}}}) = \exp\{-E_0(\rho, W_{\mathsf{BSC},\overline{\mathsf{BSC}}}^+)\},$$

and

$$\exp\{-E_0(\rho, W_{1,2}^+)\} = \mathbb{E}\left[H_{g^{-1}(\rho,T_2),\rho}(T_1)\right] \geq 2^{-\rho} + P(Z_{\mathsf{BEC}} = 0)P(Z_{\overline{\mathsf{BEC}}} = 0)\left(1 - 2^{-\rho}\right)$$
$$= \exp\{-E_0(\rho, W_{\mathsf{BEC},\overline{\mathsf{BEC}}}^+)\}.$$

Similarly, the convexity of the function $H_{z,\rho}(t)$ with respect to t for $\rho \in [1,2]$ leads to the reverse inequalities in (18). ∎

C. Special ρ Values

In Theorem 1, we have shown that among all B-DMC's W of fixed $E_0(\rho, W)$, the binary erasure channel's minus transformation results in a lower bound to any $E_0(\rho, W^-)$ and the binary symmetric channel's one in an upper bound to any $E_0(\rho, W^-)$. For the plus transformation, a similar extremality property holds except the difference that the result breaks into two parts depending on the value of the parameter ρ: While the binary erasure and binary symmetric channels appear on opposite sides of the inequalities for $E_0(\rho, W^-)$ and $E_0(\rho, W^+)$ when $\rho \in [0,1] \cup [2,\infty]$, they appear on the same side when $\rho \in [1,2]$. Using these results, we identify in this section some special cases of ρ values to recover known, and discover new results.

1) $\rho = 0$, Symmetric capacity: In [3], it is shown that the symmetric capacity is preserved under the basic polarization transformations. This property holds regardless of whether the combined channels are identical or not, as it is a consequence of the chain rule for mutual information. Namely, the channels satisfy:

$$2I(W) = I(W^-) + I(W^+).$$

This relation implies the process attached to the symmetric capacities of the synthesized channels is a bounded martingale, hence converges almost surely.

Corollary 1: Under the assumptions as Theorem 1 with $W_1 = W_2 = W$, we have

$$E_0(\rho, W_{\mathsf{BSC}}^+) - E_0(\rho, W_{\mathsf{BSC}}^-) \le E_0(\rho, W^+) - E_0(\rho, W^-) \le E_0(\rho, W_{\mathsf{BEC}}^+) - E_0(\rho, W_{\mathsf{BEC}}^-).$$

for $\rho \in [0, 1]$

Corollary 1 shows that amongst channels W with a given value of $E_0(\rho, W)$ for a given ρ the BEC and BSC are the most and least polarizing under Arıkan's polar transformations in the sense that their polar transforms W^+ and W^- has the largest and smallest difference in their E_0 values. Dividing all sides of the inequality above by ρ and taking the limit as $\rho \to 0$, we see that among channels of a given symmetric capacity, the BEC and BSC are extremal with respect to the polarization transformations, in the sense that

$$I(W_{\mathsf{BSC}}^+) - I(W_{\mathsf{BSC}}^-) \le I(W^+) - I(W^-) \le I(W_{\mathsf{BEC}}^+) - I(W_{\mathsf{BEC}}^-).$$

This is a known argument proving the convergence is to the extremes of the $[0, 1]$ interval. The preservation property of the symmetric capacities holds regardless of whether the combined channels are identical or not, as it is a consequence of the chain rule for mutual information. Namely, the channels satisfy:

$$I(W_1) + I(W_2) = I(W_{1,2}^-) + I(W_{1,2}^+),$$

and Theorem 1 can be used to show the convergence is also to the extremes values $\{0, 1\}$ of the corresponding bounded martingale process.

Remark 3: These inequalities for the symmetric capacities can also be obtained by the results on the extremes of information combining [7], together with the fact that symmetric capacity is preserved under the polarization transformations [3].

2) $\rho = 1$, Cut-off rate, Bhatthacharyya parameter: Another result of [3] can be recovered by letting $\rho = 1$. In this case, Theorem 1 implies channels having equal cut-off rates satisfy

$$E_0(1, W_{\mathsf{BEC}}^-) \le E_0(1, W^-) \le E_0(1, W_{\mathsf{BSC}}^-),$$

$$E_0(1, W_{\mathsf{BSC}}^+) = E_0(1, W^+) = E_0(1, W_{\mathsf{BEC}}^+).$$

Moreover, by the definition in Equation (3), the extremalities for the Bhattacharyya parameter are also obtained. Indeed, we know $Z(W^+) = Z(W)^2$ by [3].

3) $\rho = 2$: A previously unknown result is found by taking $\rho = 2$ in the theorem. Similar to the case $\rho = 1$, we observe the E_0 parameter of the channels $W^+, W_{\mathsf{BEC}}^+, W_{\mathsf{BSC}}^+$ are equal to each other.

D. Generalizations of the Bhatthacharyya parameter

In this section, we discuss a generalization to the definition of the Bhattacharyya parameter. We propose an extension motivated by the E_0 parameter of BECs. Given a BEC W_{BEC} with erasure probability ϵ_{bec}, we have

$$\epsilon_{\mathsf{bec}} = \frac{2^\rho 2^{-E_0(\rho, W_{\mathsf{BEC}})} - 1}{2^\rho - 1}.$$

We also know the Bhattacharyya parameter of a binary erasure channel satisfies $Z(W_{\mathsf{BEC}}) = \epsilon_{\mathsf{bec}}$. This parameter provides tighter bounds than $E_0(1, W)$ in [3], and is used in the subsequent analysis. This gives the idea to define a similar quantity to $Z(W)$, referred as $Z(\rho, W)$, which reflects the dependence on the value of ρ

$$Z(\rho, W) = \frac{2^\rho 2^{-E_0(\rho, W)} - 1}{2^\rho - 1}.$$

Using the results we derived in the previous section, the next Corollary shows how $Z(\rho, W)$ is affected by the basic channel transformations.

Corollary 2: Given a B-DMC W, for any fixed value of $\rho \geq 0$, we define a binary symmetric channel W_{BSC}, and a binary erasure channel W_{BEC} through the equality

$$Z(\rho, W) = Z(\rho, W_{\mathsf{BEC}}) = Z(\rho, W_{\mathsf{BSC}})$$

Then for the W^- and W^+ polar transformations, we have

$$Z(\rho, W_{\mathsf{BSC}}^-) \leq Z(\rho, W^-) \leq Z(\rho, W_{\mathsf{BEC}}^-) = 2Z(\rho, W_{\mathsf{BEC}}) - Z(\rho, W_{BEC})^2,$$

$$Z(\rho, W_{\mathsf{BEC}})^2 = Z(\rho, W_{\mathsf{BEC}}^+) \leq Z(\rho, W^+) \leq Z(\rho, W_{\mathsf{BSC}}^+) \qquad \forall\, \rho \in [0, 1] \cup [2, \infty],$$

$$Z(\rho, W_{\mathsf{BSC}}^+) \leq Z(\rho, W^+) \leq Z(\rho, W_{\mathsf{BEC}}^+) = Z(\rho, W_{\mathsf{BEC}})^2 \ \forall\, \rho \in [1, 2]. \quad (22)$$

III. Conclusions

The extremality of the BEC and BSC for polar transforms can be interpreted in the context of information combining. Theorem 1 shows that even if we change the measure of information from the customary mutual information to E_0 the channels BEC and BSC still remain extremal.

The results of the theorem also show the $\rho = 1, 2$ values share a common property: One can recover the value of the parameter $E_0(\rho, W^+)$ from the value of $E_0(\rho, W)$ without necessarily knowing the particular channel W. Finally, the extremality results of the theorem open up the possibility to apply the theory of channel polarization to combining arbitrary B-DMCs, the details of which will further be investigated in a future work.

IV. ACKNOWLEDGMENT

APPENDICES

In these appendices, we prove in part A Lemma 5, and in part B Lemma 6. For the proofs, we need the following lemma.

Lemma 7: The function $g(\rho, z)$ defined as

$$g(\rho, z) = \left(\frac{1}{2}(1+z)^{\frac{1}{1+\rho}} + \frac{1}{2}(1-z)^{\frac{1}{1+\rho}} \right)^{1+\rho},$$

for $z \in [0, 1]$, and $\rho \in \mathbf{R} \setminus \{-1\}$, is a concave non-increasing function in z for $\rho \in (-\infty, -1) \cup [0, \infty)$, and a convex non-decreasing function in z for $\rho \in (-1, 0]$.

Proof: Taking the first derivative with respect to z, we get

$$\frac{\partial g(\rho, z)}{\partial z} = \left(\frac{1}{2}(1+z)^{\frac{1}{1+\rho}} + \frac{1}{2}(1-z)^{\frac{1}{1+\rho}} \right)^{\rho} \left(\frac{1}{2}(1+z)^{\frac{-\rho}{1+\rho}} - \frac{1}{2}(1-z)^{\frac{-\rho}{1+\rho}} \right)$$

$$= \underbrace{\left(\frac{1}{2} \right)^{1+\rho} \left(1 + \left(\frac{1-z}{1+z} \right)^{\frac{1}{1+\rho}} \right)^{\rho}}_{\geq 0} \left(1 - \left(\frac{1-z}{1+z} \right)^{\frac{-\rho}{1+\rho}} \right). \quad (23)$$

As we have

$$\frac{1-z}{1+z} \leq 1,$$

for $\forall z \in [0, 1]$, the monotonicity claims follow by noting that when $\rho \in (-\infty, -1) \cup [0, \infty)$:

$$\frac{\rho}{1+\rho} \geq 0 \quad \Rightarrow \quad \left(1 - \left(\frac{1-z}{1+z} \right)^{\frac{-\rho}{1+\rho}} \right) \leq 0 \quad \Rightarrow \quad \frac{\partial g(\rho, z)}{\partial z} \leq 0,$$

and when $\rho \in (-1, 0]$:

$$\frac{\rho}{1+\rho} \leq 0 \quad \Rightarrow \quad \left(1 - \left(\frac{1-z}{1+z} \right)^{\frac{-\rho}{1+\rho}} \right) \geq 0 \quad \Rightarrow \quad \frac{\partial g(\rho, z)}{\partial z} \geq 0.$$

Taking the second derivative with respect to z, we get

$$\frac{\partial^2 g(\rho, z)}{\partial z^2} = -\frac{\rho}{1+\rho} \underbrace{\left(1-z^2\right)^{\frac{1}{1+\rho}-2} \left(\frac{1}{2}(1+z)^{\frac{1}{1+\rho}} + \frac{1}{2}(1-z)^{\frac{1}{1+\rho}}\right)^{-1+\rho}}_{\geq 0}.$$

The convexity claims follow once more by inspecting the sign of $\frac{\rho}{1+\rho}$ in different intervals, i.e. when $\rho \in (-\infty, -1) \cup [0, \infty)$:

$$\frac{\rho}{1+\rho} \geq 0 \quad \Rightarrow \quad \frac{\partial^2 g(\rho, z)}{\partial z^2} \leq 0,$$

and when $\rho \in (-1, 0]$:

$$\frac{\rho}{1+\rho} \leq 0 \quad \Rightarrow \quad \frac{\partial^2 g(\rho, z)}{\partial z^2} \geq 0.$$

∎

Appendix A

Proof of Lemma 5: We prove that the function $F_{z,\rho}(t) = g(\rho, zg^{-1}(\rho, t))$ defined in Equation 19 is convex with respect to the variable t for fixed $\rho \geq 0$ and $z \in [0, 1]$ values. Taking the first derivative with respect to t, we obtain

$$\begin{aligned}
\frac{\partial F_{z,\rho}(t)}{\partial t} &= \frac{\partial}{\partial t} g(\rho, zg^{-1}(\rho, t)) \\
&= \frac{g'(\rho, zg^{-1}(\rho_1, t))}{g'(\rho_1, g^{-1}(\rho_1, t))} z.
\end{aligned}$$

We define $u = g^{-1}(\rho, t)$. Since $g(\rho, u)$ is a non-increasing function in u when $\rho \geq 0$ by Lemma 7, so is $g^{-1}(\rho, t)$ in t. Hence we can check the convexity of $F_{z,\rho}(t)$ with respect to the variable t, from the monotonicity with respect to u of the following expression:

$$z\frac{g'(\rho_2, zu)}{g'(\rho_1, u)}. \tag{24}$$

To simplify notation, we define

$$f(u) = \frac{1-u}{1+u} \tag{25}$$

$$\alpha(\rho, u) = (1 + f(u)^{\frac{1}{1+\rho}})^\rho \geq 0 \tag{26}$$

$$\beta(\rho, u) = (1 - f(u)^{\frac{-\rho}{1+\rho}}) \leq 0 \tag{27}$$

Then, by equation (23)

$$\frac{\partial g(\rho, u)}{\partial u} = \left(\frac{1}{2}\right)^{1+\rho} \alpha(\rho, u)\beta(\rho, u)$$

Similarly,

$$\frac{\partial g(\rho, zu)}{\partial u} = zg'(\rho, zu)$$
$$= \left(\frac{1}{2}\right)^{1+\rho} z\alpha(\rho, zu)\beta(\rho, zu),$$

and (24) is given by

$$z\frac{g'(\rho_2, zu)}{g'(\rho_1, u)} = z\frac{\alpha(\rho, zu)\beta(\rho, zu)}{\alpha(\rho, u)\beta(\rho, u)}. \tag{28}$$

Now taking the derivative of (28) with respect to u, we get

$$\frac{\partial}{\partial u} \; z\frac{\alpha(\rho, zu)\beta(\rho, zu)}{\alpha(\rho, u)\beta(\rho, u)}$$
$$= z\underbrace{\frac{\alpha(\rho, zu)\beta(\rho, zu)}{\alpha(\rho, u)\beta(\rho, u)}}_{\geq 0}$$
$$\left(\frac{\partial\alpha(\rho, zu)/\partial u}{\alpha(\rho, zu)} + \frac{\partial\beta(\rho, zu)/\partial u}{\beta(\rho, zu)} - \frac{\partial\alpha(\rho, u)/\partial u}{\alpha(\rho, u)} - \frac{\partial\beta(\rho, u)/\partial u}{\beta(\rho, u)}\right) \tag{29}$$

We can see that the sign of the expression inside the parenthesis in (29) will determine the monotonicity in u of the expression in (28). At this point, we note that

$$\frac{\partial\alpha(\rho, u)/\partial u}{\alpha(\rho, u)} + \frac{\partial\beta(\rho, u)/\partial u}{\beta(\rho, u)} = \left(\frac{\partial\alpha(\rho, zu)/\partial u}{\alpha(\rho, zu)} + \frac{\partial\beta(\rho, zu)/\partial u}{\beta(\rho, zu)}\right)\Big|_{z=1} \tag{30}$$

Moreover, we claim that the expression inside the parenthesis in the RHS of (30) is non-decreasing in z. As a consequence, $F_{z,\rho}(t)$ is a concave function in $u = g^{-1}(\rho, t)$. Since u is decreasing in t, we have

$$\frac{\partial^2 F_{z,\rho}(t)}{\partial t^2} = \underbrace{\frac{\partial}{\partial u}\left(z\frac{g'(\rho_2, zu)}{g'(\rho_1, u)}\right)}_{\leq 0}\underbrace{\frac{\partial u}{\partial t}}_{\leq 0} \geq 0.$$

We conclude that $F_{z,\rho}(t)$ is a convex function with respect to variable t.

In the rest of the appendix, we prove our claim. We have,

$$\frac{\partial\alpha(\rho, zu)}{\partial u} = \frac{\rho}{1+\rho}zf'(zu)f(zu)^{\frac{-\rho}{1+\rho}}(1 + f(zu)^{\frac{1}{1+\rho}})^{\rho-1} \tag{31}$$

$$\frac{\partial\beta(\rho, zu)}{\partial u} = \frac{\rho}{1+\rho}zf'(zu)f(zu)^{\frac{-\rho}{1+\rho}-1} \tag{32}$$

where

$$f'(u) = \frac{\partial f(u)}{\partial u} = \frac{-2}{(1+u)^2}.$$

Hence,

$$\frac{\partial\alpha(\rho,zu)/\partial u}{\alpha(\rho,zu)}+\frac{\partial\beta(\rho,zu)/\partial u}{\beta(\rho,zu)}$$

$$=\frac{\rho}{1+\rho}f(zu)^{\frac{-\rho}{1+\rho}-1}zf'(zu)\left(\frac{f(zu)}{1+f(zu)^{\frac{1}{1+\rho}}}+\frac{1}{1-f(zu)^{\frac{-\rho}{1+\rho}}}\right)$$

$$=\frac{\rho}{1+\rho}f(zu)^{\frac{-\rho}{1+\rho}-1}zf'(zu)\left(\frac{f(zu)-f(zu)^{\frac{1}{1+\rho}}+1+f(zu)^{\frac{1}{1+\rho}}}{(1+f(zu)^{\frac{1}{1+\rho}})(1-f(zu)^{\frac{-\rho}{1+\rho}})}\right)$$

$$=\frac{\rho}{1+\rho}f(zu)^{\frac{-\rho}{1+\rho}-1}zf'(zu)(1+f(zu))(1+f(zu)^{\frac{1}{1+\rho}})^{-1}(1-f(zu)^{\frac{-\rho}{1+\rho}})^{-1}$$

$$=\frac{\rho}{1+\rho}zf'(zu)(1+f(zu)^{-1})(1+f(zu)^{\frac{1}{1+\rho}})^{-1}(f(zu)^{\frac{\rho}{1+\rho}}-1)^{-1}$$

$$=\frac{\rho}{1+\rho}\frac{-4z}{(1+zu)^2(1-zu)}\left(1+\left(\frac{1-zu}{1+zu}\right)^{\frac{1}{1+\rho}}\right)^{-1}\left(-1+\left(\frac{1-zu}{1+zu}\right)^{\frac{\rho}{1+\rho}}\right)^{-1}$$

$$=\frac{4\rho}{1+\rho}\left(\underbrace{\frac{1-z^2u^2}{z}\left((1+zu)^{\frac{\rho}{1+\rho}}-(1-zu)^{\frac{\rho}{1+\rho}}\right)}_{\text{Part 2}}\underbrace{\left((1+zu)^{\frac{1}{1+\rho}}+(1-zu)^{\frac{1}{1+\rho}}\right)}_{\text{Part 1}}\right)^{-1}$$

We consider the expressions labeled as Part 1 and Part 2 separately. Note that both are positive valued. In addition, we will show that both are decreasing in z. As a result, we deduce

$$\frac{\partial}{\partial z}\left(\frac{(1-z^2u^2)}{z}\left((1+zu)^{\frac{1}{1+\rho}}+(1-zu)^{\frac{1}{1+\rho}}\right)\left((1+uz)^{\frac{\rho}{1+\rho}}-(1-uz)^{\frac{\rho}{1+\rho}}\right)\right)\leq 0$$

$$\frac{\partial}{\partial z}\left(\frac{(1-z^2u^2)}{z}\left((1+zu)^{\frac{1}{1+\rho}}+(1-zu)^{\frac{1}{1+\rho}}\right)\left((1+uz)^{\frac{\rho}{1+\rho}}-(1-uz)^{\frac{\rho}{1+\rho}}\right)\right)^{-1}\geq 0$$

which is proves our claim.

For Part 1, we get

$$\frac{\partial}{\partial z}\left((1+zu)^{\frac{1}{1+\rho}}+(1-zu)^{\frac{1}{1+\rho}}\right)$$

$$=\frac{u\left((1+uz)^{\frac{-\rho}{1+\rho}}-(1-uz)^{\frac{-\rho}{1+\rho}}\right)}{1+\rho}\leq 0$$

For Part 2, we have

$$\frac{\partial}{\partial z}\left(\frac{(1-u^2z^2)}{z}\left((1+uz)^{\frac{\rho}{1+\rho}}-(1-uz)^{\frac{\rho}{1+\rho}}\right)\right)$$

$$=\frac{1}{z^2}\frac{\rho uz\left(1-u^2z^2\right)\left((1+uz)^{\frac{\rho}{1+\rho}-1}+(1-uz)^{\frac{\rho}{1+\rho}-1}\right)}{1+\rho}$$

$$+\frac{1}{z^2}\left(1+u^2z^2\right)\left(-(1+uz)^{\frac{\rho}{1+\rho}}+(1-uz)^{\frac{\rho}{1+\rho}}\right)$$

$$=\frac{1}{z^2}\left(\ (1+uz)^{\frac{\rho}{1+\rho}}\left(\frac{\rho}{1+\rho}uz\left(1-uz\right)-\left(1+u^2z^2\right)\right)\right.$$

$$\left.+\left(1-uz\right)^{\frac{\rho}{1+\rho}}\left(\frac{\rho}{1+\rho}uz\left(1+uz\right)+\left(1+u^2z^2\right)\right)\right)$$

$$=\frac{1}{z^2}\left(\ -(1+x)^k\left((k+1)\,x^2-kx+1\right)+(1-x)^k\left((k+1)\,x^2+kx+1\right)\right)$$

$$=\frac{1}{z^2}\left(\ -f_1(x,k)+f_2(x,k)\right) \tag{33}$$

where $k=\frac{\rho}{1+\rho}\in[0,1)$, $x=uz\in[0,1]$, and

$$f_1(x,k)=(1+x)^k\left((k+1)\,x^2-kx+1\right), \tag{34}$$

$$f_2(x,k)=(1-x)^k\left((k+1)\,x^2+kx+1\right). \tag{35}$$

We will show that $f_1(x,k)\geq f_2(x,k)$ holds for $x\in[0,1]$, and for $k\in[0,1)$. Since $f_1(x,k),f_2(x,k)\geq 0$, this is equivalent to showing that $\log\dfrac{f_1(x,k)}{f_2(x,k)}\geq 0$ holds. We have

$$\log\frac{f_1(x,k)}{f_2(x,k)}=k\log\frac{1+x}{1-x}+\log\left((k+1)\,x^2-kx+1\right)-\log\left((k+1)\,x^2+kx+1\right).$$

We immediately observe that when $k=0$ we have the above sum equals to 0. Now, we will show that

$$\frac{\partial}{\partial k}\log\frac{f_1(x,k)}{f_2(x,k)}\geq 0.$$

Hence, this will prove our claim that $f_1(x,k)\geq f_2(x,k)$ holds.

Taking the first derivative with respect to k, we have

$$\frac{\partial}{\partial k}\log\frac{f_1(x,k)}{f_2(x,k)}=\log\frac{1+x}{1-x}-\frac{2x\left(1+x^2\right)}{\left(1+(k+1)x^2\right)^2-(kx)^2}.$$

So, we will be done if

$$\log\frac{1+x}{1-x}\geq 2x\left(1+x^2\right)\max_{k\in[0,1)}\frac{1}{\left(1+(k+1)x^2\right)^2-(kx)^2}.$$

One can easily check that the expression in the denominator $(1 + (k+1)x^2)^2 - (kx)^2$ is non-decreasing in $k \in [0, 1)$, hence the reciprocal is non-increasing in k. As a result, the maximum is attained at $k = 0$. Therefore, we only have to prove that

$$\log \frac{1+x}{1-x} \geq \frac{2x(1+x^2)}{(1+x^2)^2} = \frac{2x}{(1+x^2)}$$

holds. But, we have

$$\log \frac{1+x}{1-x} = 2x \left(1 + \frac{1}{3}x^2 + \frac{1}{5}x^4 + \frac{1}{7}x^6 + \dots\right) \geq 2x \geq \frac{2x}{(1+x^2)}.$$

So, $-f_1(x, k) + f_2(x, k) \leq 0$ holds for $k \in [0, 1)$ and $x \in [0, 1]$. Consequently, Part 2 is also decreasing in z. This proves our claim that the RHS of (30) is non-decreasing in z. ∎

Appendix B

Proof of Lemma 6: In this Appendix, we show that the function $H_{z,\rho}(t) = h(\rho, g^{-1}(\rho, t), z)$ defined in Equation 6 is concave with respect to the variable t when $\rho \in [0, 1] \cup [2, \infty]$, and convex otherwise when $\rho \in [1, 2]$, for any fixed $z \in [0, 1]$, and $\rho \geq 0$.

Taking the first derivative with respect to t, we get

$$\frac{\partial}{\partial t} H_{z,\rho}(t) = \frac{h'(\rho, g^{-1}(\rho, t), z)}{g'(\rho, g^{-1}(\rho, k))}.$$

As we did in Appendix A, we define $u = g^{-1}(\rho_1, t)$. Since $g(\rho, u)$ is a non-increasing function in u by Lemma 7, so is $g^{-1}(\rho_1, t)$ in t. Hence we can check the concavity of $H_{z,\rho_1}(t)$ with respect to variable t, by verifying that

$$\frac{h'(\rho, u, z)}{g'(\rho, u)}$$

is non-decreasing in u. So, we check that

$$\frac{\partial}{\partial u} \left(\frac{h'(\rho, u, z)}{g'(\rho, u)} \right) = \frac{h''(\rho, u, z)g'(\rho, u) - h'(\rho, u, z)g''(\rho, u)}{g'(\rho, u)^2} \geq 0.$$

Since the denominator is always positive, we only need to show that

$$h''(\rho, u, z)g'(\rho, u) - h'(\rho, u, z)g''(\rho, u) \geq 0. \tag{36}$$

Moreover, we observe that $h(\rho, u, 0) = g(\rho, u)$. So, we can equivalently show the following relation holds:

$$\frac{h''(\rho, u, z)}{h'(\rho, u, z)} \geq \frac{h''(\rho, u, 0)}{h'(\rho, u, 0)}. \tag{37}$$

We first apply the transformations

$$u = \tanh(k), \quad z = \tanh(w)$$

where $k, w \in [0, \infty)$. For shorthand notation, let $h(\rho, \tanh(k), \tanh(w)) \triangleq \tilde{h}(\rho, k, w)$. Using these, we obtain

$$\tilde{h}(\rho, k, w) = \frac{\cosh(\frac{1}{1+\rho}(k+w))^{1+\rho} + \cosh(\frac{1}{1+\rho}(k-w))^{1+\rho}}{2\cosh(k)\cosh(w)}.$$

Then,

$$\frac{\frac{\partial h(\rho, k, w)}{\partial k^2}}{\frac{\partial h(\rho, k, w)}{\partial k}} = -2\tanh(k) + \frac{\rho}{1+\rho}\cosh(k) \times$$

$$\left[\frac{\cosh(\frac{1}{1+\rho}(k+w))^{\rho-1} + \cosh(\frac{1}{1+\rho}(k-w))^{\rho-1}}{\cosh(\frac{1}{1+\rho}(k+w))^{\rho}\sinh(\frac{\rho}{1+\rho}k - \frac{1}{1+\rho}w) + \cosh(\frac{1}{1+\rho}(k-w))^{\rho}\sinh(\frac{\rho}{1+\rho}k + \frac{1}{1+\rho}w)} \right]. \quad (38)$$

We note that the additive term $-2\tanh(k)$, and the non-negative multiplicative factor $\frac{\rho}{1+\rho}\cosh(k)$ do not depend on w. Hence, we only need to show the term inside the parenthesis is smallest when evaluated at $w = 0$. For this purpose, we define the transformations

$$a = \frac{k+w}{1+\rho}, \quad b = \frac{k-w}{1+\rho}$$

such that $k = (1+\rho)\frac{a+b}{2}$, and $w = (1+\rho)\frac{a-b}{2}$. The condition $k, w \geq 0$ is equivalent to $a \geq |b|$. Using these transformations, the reciprocal of the term inside parenthesis in equation (38) becomes

$$R(\rho, a, b) = \frac{\cosh(b)^{1-\rho}\cosh(a)\sinh(\frac{a+b}{2}\rho - \frac{a-b}{2}) + \cosh(a)^{1-\rho}\cosh(b)\sinh(\frac{a+b}{2}\rho + \frac{a-b}{2})}{\cosh(a)^{1-\rho} + \cosh(b)^{1-\rho}}.$$

Therefore, the inequality given in (37) will hold iff

$$R(\rho, a, b) \leq R(\rho, \frac{a+b}{2}, \frac{a+b}{2}) = \cosh(\frac{a+b}{2})\sinh(\frac{a+b}{2}\rho). \quad (39)$$

We define

$$f(\rho, a, b) \triangleq \cosh(\frac{a+b}{2})\sinh(\rho\frac{a+b}{2})\left[\cosh(a)^{1-\rho} + \cosh(b)^{1-\rho}\right]$$

$$- \cosh(a)^{1-\rho}\cosh(b)\sinh(\rho\frac{a+b}{2} + \frac{a-b}{2}) - \cosh(b)^{1-\rho}\cosh(a)\sinh(\rho\frac{a+b}{2} - \frac{a-b}{2}).$$

We note that $f(\rho, a, b) \geq 0$ is equivalent to the inequality (39), which in turn is equivalent to the inequality (37).

After simplifications, the function reduces to the following form:

$$f(\rho, a, b) = \sinh(\frac{a-b}{2}) J(\rho, a, b)$$

where

$$J(\rho, a, b) \triangleq \cosh(b)^{1-\rho} \cosh(a - \rho\frac{a+b}{2}) - \cosh(a)^{1-\rho} \cosh(b - \rho\frac{a+b}{2}).$$

Since for $a \geq |b|$, we have

$$\sinh(\frac{a-b}{2}) \geq 0$$

we only need to show that $J(\rho, a, b) \geq 0$.

We introduce the variables k', and w' using $a = k' + w'$, and $b = k' - w'$ where $k', w' \in [0, \infty)$. Then, we get

$$J(\rho, k'+w', k'-w') = \cosh(k'-w')^{1-\rho} \cosh(k'-\rho k'+w') - \cosh(k'-\rho k'-w') \cosh(k'+w')^{1-\rho}.$$

We note that $J(\rho, k' + w', k' - w')\big|_{k'=0} = 0$. Moreover, $J(\rho, k' + w', k' - w')$ is increasing in the variable k': taking the first derivative with respect to k', we get

$$\frac{\partial}{\partial k'} J(\rho, k'+w', k'-w') = (1-\rho)\left[\cosh(k'-w')^{-\rho} - \cosh(k'+w')^{-\rho}\right]\sinh((2-\rho)k') \geq 0$$

where the positivity follows from the fact that $|k' - w'| \leq |k' - w'|$, thus $\cosh(k' - w') \leq \cosh(k' + w')$, and $\cosh(k' - w')^{-\rho} \geq \cosh(k' + w')^{-\rho}$, and from the fact that $\sinh(x) \geq 0$ holds for $\forall x \geq 0$.

As a result, $J(\rho, k' + w', k' - w') \geq 0$ as required, and we have shown that the inequality given in (37) holds. This concludes the proof. ∎

REFERENCES

[1] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, NY, USA, 1968.

[2] M. Alsan. Extremality properties for Gallager's random coding exponent. *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2944–2948, 1-6 July 2012.

[3] E. Arıkan. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theor.*, 55(7):3051–3073, 2009.

[4] E. Arıkan and E. Telatar. BEC and BSC are E_0 extremal. Unpublished manuscript. July 2008.

[5] M. Alsan, E. Telatar. Submartingale Property of E_0 Under The Polarization Transformations *eprint arXiv:1207.6788* July 2012

[6] E. Telatar, Private Communications.

[7] I. Sutskover, S. Shamai, and J. Ziv. Extremes of information combining. *Information Theory, IEEE Transactions on*, 51(4):1313 – 1325, April 2005.

[8] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products*. Academic Press Inc, 1994.